

Oponentský posudek na disertační práci:

R. SABOLOVÁ: STATISTICAL INFERENCE BASED ON SADDLEPOINT APPROXIMATIONS

Předložená disertační práce se zabývá zejména použitím metody sedlového bodu na aproximaci rozdělení odhadů parametrů v modelu kvantilové regrese a na odvození vhodných testových statistik.

První kapitola přehledně popisuje aproximaci hustoty M-odhadu metodou sedlového bodu, sekce 1.2 porovnává skutečnou a aproximovanou hustotu pro některé speciální případy. V sekci 1.3 jsou popsány testy navržené v článcích Ronchetti et al (2003) a Ma & Ronchetti (2011).

Druhá kapitola se zabývá použitím metod sedlového bodu na kvantily a v modelu kvantilové regrese. Autorka nejdříve odvozuje aproximace za předpokladu známého rozdělení pozorování (parametrický přístup) a posléze popisuje i postup založený na empirické distribuční funkci residuí (neparametrický přístup). V sekci 2.3 je odvozena aproximace pro průměrné regresní kvantily (averaged regression quantiles).

Třetí kapitola obsahuje nejdůležitější výsledky, kterými jsou dva testy o hodnotě vektoru parametrů v modelu kvantilové regrese. V sekci 3.1 je popsán test založený na průměrných regresních kvantilech (za předpokladu známého rozdělení chybových členů). V sekci 3.2 je metodou sedlového bodu zkoustruována jednorozměrná testová statistika (s asymptotickým χ^2 rozdělením) podle článků Ronchetti et al (2003) a Ma & Ronchetti (2011). Podobně jako ve druhé kapitole, i zde autorka rozlišuje parametrický a neparametrický případ. V zajímavějším a prakticky použitelném neparametrickém případě (věta 10 na straně 52) však už bohužel chybí formální odvození “předpokládaného” asymptotického rozdělení (viz poznámka 7 na straně 55). Sekce 3.3 popisuje testy o podvektoru β a sekce 3.4 obsahuje simulační studii.

Ve čtvrté kapitole je odvozena aproximace rozdělení maximálně věrohodného odhadu metodou sedlového bodu založená na I-divergenci. Autorka ukazuje, že pro exponenciální rodinu rozdělení tento postup vede na explicitní vzorec. Postup je použit na skutečná data (metan) předpokládající Paretovo rozdělení.

Připomínky:

1. Značení v kapitole 2: co znamenají symboly $F()$, g , g_n , f_n a proč se význam těchto symbolů v průběhu kapitoly neustále mění?
2. Překlep v posledním vzorci na straně 42.
3. Značení v kapitole 4: co znamenají symboly Z , Z_- a Z_+ (str. 78 dole). Proč a jak se metanová data “ořezávají”?

Formálně je předložená práce zpracována velice pěkně, pouze s minimem překlepů. Ve druhé kapitole se mi však příliš nelíbí použité značení, které vypadá jako by tato kapitola byla poslepovaná ze dvou různých článků. Některé symboly použité v příkladu ve čtvrté kapitole by mohly být podrobněji vysvětleny (například $\hat{\gamma}_i$ nebo $Z+$).

Nové výsledky jsou odvozeny zejména ve třetí a ve čtvrté kapitole. U prakticky použitelné testové statistiky odvozené ve Větě 10 však (zatím?) chybí odvození asymptotického rozdělení.

Dotazy a náměty pro diskusi:

1. Proč se v simulační studii v sekci 3.4 neporovnává navržený postup i s metodou bootstrap?
2. Jak je u grafické metody testování (sekce 4.2.2, str. 74) zaručeno dodržení hladiny testu?
3. Celá práce se soustředí na aproximaci *hustoty* odhadu nebo testové statistiky. Pro testování (např. ve druhé nebo čtvrté kapitole) by však byla mnohem užitečnější distribuční funkce. Jak se vlastně z odvozené teoretické hustoty v praxi získává například p-hodnota?

Podle mého názoru se celkově jedná o velmi pěknou disertační práci, která zcela jednoznačně prokazuje předpoklady autorky k samostatné tvořivé práci.

Doc. RNDr. Zdeněk Hlávka, Ph.D.